CERIAS Tech Report 2001-142

The EM/MPM Algorithm for Segmentation of Textured Images: Analysis and Further Experimental Results

by M Comer, E Delp
Center for Education and Research
Information Assurance and Security
Purdue University, West Lafayette, IN 47907-2086

THE EM/MPM ALGORITHM FOR SEGMENTATION OF TEXTURED IMAGES: ANALYSIS AND FURTHER EXPERIMENTAL RESULTS

Mary L. Comer and Edward J. Delp
Computer Vision and Image Processing Laboratory
School of Electrical Engineering
Purdue University
West Lafayette, Indiana

Corresponding Author:
Professor Edward J. Delp
School of Electrical Engineering
1285 Electrical Engineering Building
Purdue University
West Lafayette, IN 47907-1285
Telephone: (317) 494-1740
Fax: (317) 494-0880
ace@ecn.purdue.edu

In this paper we present new results relative to the "expectation-maximization/maximization of the posterior marginals" (EM/MPM) algorithm for simultaneous parameter estimation and segmentation of textured images. The EM/MPM algorithm uses a Markov random field model for the pixel class labels and alternately approximates the MPM estimate of the pixel class labels and estimates parameters of the observed image model. The goal of the EM/MPM algorithm is to minimize the expected value of the number of misclassified pixels. We present new theoretical results in this paper which show that the algorithm can be expected to achieve this goal, to the extent that the EM estimates of the model parameters are close to the true values of the model parameters. We also present new experimental results demonstrating the performance of the EM/MPM algorithm.

EDICS: IP 1.5

This work was partially supported by a National Science Foundation Graduate Fellowship.

1 INTRODUCTION

This paper addresses the problem of segmenting a textured image. In the observed image, there are a number of regions, corresponding to different objects or different textures. Each pixel in the image must be assigned to one of a finite number of classes depending on statistical properties of the pixel and its neighbors. The individual pixel classifications, or labels, form a two-dimensional field, with the same dimensions as the observed image, in which the value at a given spatial location reflects the class to which the corresponding pixel in the observed image belongs. This two-dimensional field containing the individual pixel classifications will be referred to as the label field; the elements of the label field will be referred to as the class labels. The label field is unknown and must be estimated from the observed image.

The algorithm presented in this paper uses a statistical approach to segment textured images. Statistical segmentation schemes generally segment an image by optimizing some criterion. Several algorithms which approximate the maximum *a posteriori* (MAP) estimate of the label field given the observed image have been proposed [1, 2, 3]. Another criterion which has been used is the minimization of the expected value of the number of misclassified pixels. The estimate which optimizes this criterion is known as the "maximizer of the posterior marginals" (MPM) estimate. It has been shown that the MPM estimation criterion is more appropriate for image segmentation than the MAP criterion [4]. This is because the MAP estimate assigns the same cost to every incorrect segmentation, regardless of the number of pixels at which the incorrect segmentation differs from the true segmentation, whereas the MPM estimate assigns a cost to an incorrect segmentation based on the number of incorrectly classified pixels in that segmentation. Unfortunately, as is the case with the MAP estimate, it is computationally infeasible to compute the MPM estimate exactly. A stochastic algorithm for approximating the MPM estimate of the label field was proposed in [4]. However, this algorithm assumes that the values of all parameters for the observed image and label field models are known *a priori*. If some of these model parameters are unknown, the algorithm in [4] cannot be used.

The EM/MPM algorithm, a stochastic algorithm which combines the EM algorithm for parameter estimation with the MPM algorithm for segmentation, was proposed to address this problem [5]. The same algorithm was also proposed in [6], along with a deterministic scheme that also

approximates the MPM estimate of the label field and the EM estimates of model parameters[1]. The EM/MPM algorithm, and its deterministic counterpart from [6], estimate parameters at each stage of the algorithm using the current estimates of the marginal conditional probabilities of the class labels. This is referred to as a "soft-decision" scheme in [6], in contrast to "hard-decision" schemes which use the current segmentation to estimate parameters at each stage of the algorithm [2, 3, 7]. The soft-decision approach was shown in [6] to provide better results than the hard-decision approach.

The goal of the EM/MPM algorithm is to minimize the expected value of the number of misclassified pixels. As, shown in [4], this is equivalent to maximizing the marginal probabilities of the class labels. Since these probabilities are unknown, the maximization is performed on estimates of the probabilities. It is important to consider how close these estimates of the class label probabilities are to the true values. In this paper we present new theoretical results which address this issue. We show two important results. First, we show that the estimates of the marginal probabilities of the class labels obtained during a given stage of the EM/MPM procedure converge with probability 1 to the true values of the class label probabilities, given the estimates of the model parameters obtained during the previous stage. Second, we show that the parameter estimates resulting from the EM/MPM procedure can be made arbitrarily close to the EM estimates of the parameters with probability 1, if a sufficient number of iterations is performed.

These two results are significant because they imply that the algorithm will eventually converge to the segmentation which minimizes the expected number of misclassified pixels, as long as the EM estimates of the model parameters are close to the true values of the model parameters.

In addition to the theoretical analysis, we present experimental results comparing the performance of the EM/MPM algorithm to the deterministic EM/MPM algorithm.

In Section 2 the models used for the label field and the observed image are described. In Section 3 we describe the MPM segmentation algorithm for the case in which the values of all model parameters are known, and the EM algorithm for the case in which the values of the marginal conditional probabilities of the class labels are known. The EM/MPM algorithm is described in Section 4. Section 5 details the new theoretical results, and Section 6 contains experimental results.

[1] In this paper the stochastic EM/MPM algorithm will be referred to simply as the "EM/MPM algorithm" and the deterministic algorithm proposed in [6] will be referred to as the "deterministic EM/MPM algorithm".

2 IMAGE MODELS

In this paper the label field will be denoted $\mathbf{X}$ and the observed image will be denoted $\mathbf{Y}$. The element in $\mathbf{X}$ at spatial location $s \in S$, where S is the rectangular pixel lattice on which $\mathbf{X}$ and $\mathbf{Y}$ are defined, is the random variable denoted by X_s. This notation is also used for $\mathbf{Y}$. Throughout the paper, $\mathbf{x} = (x_1, x_2, \ldots, x_N)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_N)$, where N is the total number of pixels in S, will represent sample realizations of $\mathbf{X} = (X_1, X_2, \ldots, X_N)$ and $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_N)$. The space of all possible realizations of $\mathbf{X}$ will be denoted $\Omega_{\mathbf{x}}$ and the space of possible realizations of $\mathbf{Y}$ will be denoted $\Omega_{\mathbf{y}}$.

2.1 Markov Random Field Model

In this section we define the MRF model used for the label field in this paper. Interested readers should see [8, 9] for a detailed discussion of MRF models.

Before describing the MRF model used for $\mathbf{X}$, we must define the concept of a clique. A collection $\mathcal{G} = \{\mathcal{G}_s \subseteq S, s \in S\}$ is a neighborhood system for S if, for every pixel $s \in S$, (i) $s \notin \mathcal{G}_s$ and (ii) $s \in \mathcal{G}_r \iff r \in \mathcal{G}_s$, for any $r \in S$. The elements of the set $\mathcal{G}_s$ are the neighbors of spatial location s. A set of pixels $C \subseteq S$ is a clique if, for any pixels $s, r \in C$, $s \in \mathcal{G}_r$. Thus, the collection of all cliques, which we shall denote as $\mathcal{C}$, is induced by the neighborhood system.

The collection of cliques $\mathcal{C}$ used in this paper will include all pairs of spatially horizontally or vertically adjacent pixels, plus all single pixels. The probability mass function of $\mathbf{X}$ is assumed to have the form

$$p_{\mathbf{X}}(\mathbf{x}) = \frac{1}{z} \exp\left(- \sum_{\{r,s\} \in \mathcal{C}} \beta t(x_r, x_s) - \sum_{\{r\} \in \mathcal{C}} \gamma_{x_r} \right) \tag{1}$$

where

$$t(x_r, x_s) = \begin{cases} 0 & \text{if } x_r = x_s \\ 1 & \text{if } x_r \neq x_s \end{cases} \tag{2}$$

The parameter β is known as the spatial interaction parameter, and $\{\gamma_k\}$ is a set of model parameters for single-pixel cliques. This model is similar to MRF models previously used for segmentation [2, 7, 3]. For every pixel $s \in S$, the set of values which the random variable X_s can take is $\{1, 2, \ldots, L\}$, where L is the number of different classes, or textures, in the image. This means that $\Omega_{\mathbf{x}} = \{\mathbf{x} : x_s \in \{1, 2, \ldots, L\} \forall s \in S\}$. It will be assumed throughout this paper that L is known *a*

priori.

We assume that the value of the spatial interaction parameter is known *a priori*. This assumption is often used in MRF-based segmentation schemes [10, 7, 6]. We have found experimentally that the optimal value of β is not highly image-dependent, and that the performance of the algorithm remains fixed over a relatively large range of values of β.

The parameter γ_k can be viewed as a cost parameter for class k. If, for a given k, γ_k is high, then class k is less likely to occur than classes with lower costs. For applications in which there is *a priori* information about the relative sizes of the various classes, the parameters $\{\gamma_k\}$ can be selected to incorporate this information into the label field model. In the absence of such *a priori* information, γ_k will be assumed to be zero for every k.

2.2 Model for Observed Image

We also need a statistical model for the observed image. We will assume that the random variables $Y_1, Y_2, \ldots, Y_N$ are conditionally independent given the pixel label field $\mathbf{X}$. We will also assume that the conditional probability density function of Y_r given $\mathbf{X}$ depends only on the value of $\mathbf{X}$ at pixel location r. Using these two assumptions, the conditional probability density function of $\mathbf{Y}$ given $\mathbf{X}$ can be written as

$$\begin{aligned} f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x},\boldsymbol{\theta}) &= \prod_{r=1}^{N} f_{Y_r|\mathbf{X}}(y_r|\mathbf{x},\boldsymbol{\theta}) \\ &= \prod_{r=1}^{N} f_{Y_r|X_r}(y_r|x_r,\boldsymbol{\theta}) \end{aligned} \tag{3}$$

where $\boldsymbol{\theta}$ is a non-random vector whose elements are the unknown parameters of the conditional probability density function of $\mathbf{Y}$ given $\mathbf{X}$.

We will also model Y_r as conditionally Gaussian given X_r. The mean and variance of Y_r depend on the class to which pixel r belongs. Thus, all of the random variables in $\mathbf{Y}$ which represent class i, for any $i = 1, \ldots, L$, are independent and identically distributed (*iid*) Gaussian random variables with mean μ_i and variance σ_i^2. This model for the observed image has been used in previously proposed segmentation algorithms [7, 6].

The means and variances μ_i and σ_i^2, $i = 1, \ldots, L$, are the elements of the parameter vector $\boldsymbol{\theta}$,

i.e., $\boldsymbol{\theta} = [\mu_1, \sigma_1^2, \ldots, \mu_L, \sigma_L^2]$. The space of all possible values of $\boldsymbol{\theta}$ will be denoted Ω_θ. We will assume that the elements of $\boldsymbol{\theta}$ are unknown.

Using the form of $f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})$ given by Equation 3, the conditional probability density function of $\mathbf{Y}$ given $\mathbf{X}$ is

$$f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}) = \prod_{r=1}^{N} \frac{1}{\sqrt{2\pi\sigma_{x_r}^2}} \exp\left(-\frac{(y_r - \mu_{x_r})^2}{2\sigma_{x_r}^2}\right) \tag{4}$$

We will need to obtain the conditional probability mass function of $\mathbf{X}$ given $\mathbf{Y}$ to segment the image. Using Bayes' rule and Equations 1 and 4, we have

$$p_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta}) = \frac{f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}) p_{\mathbf{X}}(\mathbf{x})}{f_{\mathbf{Y}}(\mathbf{y}|\boldsymbol{\theta})}$$

$$= \frac{1}{f_{\mathbf{Y}}(\mathbf{y}|\boldsymbol{\theta})} \left[\prod_{r=1}^{N} \frac{1}{\sqrt{2\pi\sigma_{x_r}^2}} \exp\left(-\frac{(y_r - \mu_{x_r})^2}{2\sigma_{x_r}^2}\right)\right] \left(\frac{1}{z}\right) \exp\left(-\sum_{\{r,s\}\in\mathcal{C}} \beta t(x_r, x_s)\right)$$

$$= \frac{1}{z f_{\mathbf{Y}}(\mathbf{y}|\boldsymbol{\theta})} \left[\prod_{r=1}^{N} \frac{1}{\sqrt{2\pi\sigma_{x_r}^2}}\right] \exp\left(-\sum_{r=1}^{N} \frac{(y_r - \mu_{x_r})^2}{2\sigma_{x_r}^2} - \sum_{\{r,s\}\in\mathcal{C}} \beta t(x_r, x_s)\right) \tag{5}$$

Since $f_{\mathbf{Y}}(\mathbf{y}|\boldsymbol{\theta})$ does not depend on $\mathbf{x}$, it is not considered in the optimization. It should be noted that $p_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta})$ is also a Gibbs distribution.

3 OVERVIEW OF THE MPM AND EM ALGORITHMS

The EM/MPM algorithm is based on the MPM algorithm for segmentation and the EM algorithm for parameter estimation. In this section we present overviews of the MPM and EM algorithms to provide a foundation for describing the EM/MPM algorithm in the next section.

3.1 MPM Segmentation Algorithm

In this section we assume that $\boldsymbol{\theta}$ is known and describe the MPM segmentation algorithm. For the MPM algorithm the segmentation problem is formulated as an optimization problem. The optimization criterion which is used is the minimization of the expected value of the number of misclassified pixels. As shown in [4], minimizing this expected value is equivalent to maximizing $P(X_s = k|\mathbf{Y} = \mathbf{y})$ over all $k \in \{1, 2, \ldots, L\}$, for every $s \in S$. Thus, to find the MPM estimate of

$\mathbf{X}$, it is necessary to find for each $s \in S$ the value of k which maximizes

$$\begin{aligned} P(X_s = k|\mathbf{Y} = \mathbf{y}) &= p_{X_s|\mathbf{Y}}(k|\mathbf{y}, \boldsymbol{\theta}) \\ &= \sum_{\mathbf{x} \in \Omega_{k,s}} p_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta}) \end{aligned} \tag{6}$$

where $\Omega_{k,s} = \{\mathbf{x} : x_s = k\}$. Exact computation of these marginal probability mass functions as in Equation 6 is computationally infeasible.

Marroquin, et al. presented an algorithm for approximating these marginal probabilities to obtain an approximation to the MPM estimate of a MRF [4]. This algorithm can be used to approximate $P(X_s = k|\mathbf{Y} = \mathbf{y})$ for each $s \in S$ and $k \in \{1, 2, \ldots, L\}$ as follows: Use the Gibbs sampler [10] to generate a discrete-time Markov chain $\mathbf{X}(t)$ which converges in distribution to a random field with probability mass function $p_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta})$, given by Equation 5. The marginal conditional probability mass functions $p_{X_s|\mathbf{Y}}(k|\mathbf{y}, \boldsymbol{\theta})$, which are to be maximized, are then approximated as the fraction of time the Markov Chain spends in state k at pixel s, for each k and s.

For the Markov chain $\mathbf{X}(t)$ generated using the Gibbs sampler there are L^N possible states, corresponding to the L^N elements of $\Omega_{\mathbf{X}}$. At each step only one pixel is visited, so that $\mathbf{X}(t-1)$ and $\mathbf{X}(t)$ can differ at no more than one pixel location. At time t the state of $\mathbf{X}(t)$ at pixel s is a random variable $X_s(t)$. Let $q_t \in S$ be the pixel visited at time t. Then the state of $X_{q_t}(t)$ is determined by sampling from the conditional probability mass function $p_{X_{q_t}|\mathbf{Y}, X_r, r \in \mathcal{G}_{q_t}}(k|\mathbf{y}, x_r(t-1), r \in \mathcal{G}_{q_t}, \boldsymbol{\theta})$.

If the sequence $\{q_1, q_2, q_3, \ldots\}$ contains every pixel $s \in S$ infinitely often, then for any initial configuration $\mathbf{x}(0) \in \Omega_{\mathbf{X}}$,

$$\lim_{t \to \infty} P(\mathbf{X}(t) = \mathbf{x}|\mathbf{Y} = \mathbf{y}, \mathbf{X}(0) = \mathbf{x}(0)) = p_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta}) \tag{7}$$

for every $\mathbf{x} \in \Omega_{\mathbf{X}}$ [10]. Thus, the Markov chain converges in distribution to a random field with probability mass function $p_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta})$, i.e., $p_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta})$ is the limiting distribution of the Markov chain.

To describe the approximation of the marginal conditional probability mass function at each

pixel we first define the function

$$u_{k,s}(t) = \begin{cases} 1 & \text{if } X_s(t) = k \\ 0 & \text{if } X_s(t) \neq k \end{cases} \tag{8}$$

Then, if T_s is the number of visits to pixel s made by the Gibbs sampler, then the approximations

$$p_{X_s|\mathbf{Y}}(k|\mathbf{y}, \boldsymbol{\theta}) \approx \frac{1}{T_s} \sum_{t=1}^{T_s} u_{k,s}(t) \qquad \forall k, s \tag{9}$$

provide the estimates of the values needed to obtain $\mathbf{x}_{MPM}$.

3.2 EM Algorithm for Parameter Estimation

In order to implement the Gibbs sampler, we must estimate the value of $\boldsymbol{\theta}$. We will use the EM algorithm to estimate $\boldsymbol{\theta}$. The EM algorithm has been widely used for the estimation of parameters in incomplete-data problems [11, 12, 13, 14]. In an incomplete-data problem the observed data represent only a subset of the complete set of data. There also is a set of data which is unobserved, or hidden. For example, in our formulation the observed image $\mathbf{Y}$ represents the observed data, and the label field $\mathbf{X}$ represents the hidden data.

The EM algorithm is an iterative procedure which approximates maximum-likelihood (ML) estimates. At each iteration two steps are performed: the expectation step and the maximization step. In general, if $\boldsymbol{\theta}(p)$ is the estimate of $\boldsymbol{\theta}$ at the pth iteration, then in the expectation step at iteration p the function

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}(p-1)) = E[\log f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})|\mathbf{Y} = \mathbf{y}, \boldsymbol{\theta}(p-1)] + E[\log p_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\theta})|\mathbf{Y} = \mathbf{y}, \boldsymbol{\theta}(p-1)] \tag{10}$$

is computed. Since in our formulation the probability mass function of $\mathbf{X}$ does not depend on $\boldsymbol{\theta}$, we only use the first term of Equation 10. The estimate $\boldsymbol{\theta}(p)$ is obtained in the maximization step as the value of $\boldsymbol{\theta}$ which maximizes $Q(\boldsymbol{\theta}, \boldsymbol{\theta}(p-1))$, i.e., $\boldsymbol{\theta}(p)$ satisfies

$$Q(\boldsymbol{\theta}(p), \boldsymbol{\theta}(p-1)) \geq Q(\boldsymbol{\theta}, \boldsymbol{\theta}(p-1)) \;\; \forall \boldsymbol{\theta} \in \Omega_\theta \tag{11}$$

Substituting Equation 4 into Equation 10, differentiating, setting to zero, and solving for $\boldsymbol{\theta}(p) =$

$[\mu_1(p), \sigma_1^2(p), \ldots, \mu_L(p), \sigma_L^2(p)]$ gives

$$\mu_k(p) = \frac{1}{N_k(p)} \sum_{s=1}^{N} y_s p_{X_s|\mathbf{Y}}(k|\mathbf{y}, \boldsymbol{\theta}(p-1)) \tag{12}$$

and

$$\sigma_k^2(p) = \frac{1}{N_k(p)} \sum_{s=1}^{N} (y_s - \mu_k(p))^2 p_{X_s|\mathbf{Y}}(k|\mathbf{y}, \boldsymbol{\theta}(p-1)) \tag{13}$$

where

$$N_k(p) = \sum_{s=1}^{N} p_{X_s|\mathbf{Y}}(k|\mathbf{y}, \boldsymbol{\theta}(p-1)) \tag{14}$$

for $k = 1, \ldots, L$.

Equations 12 through 14 demonstrate the difficulty in using the EM algorithm with a MRF model for the pixel label field. The conditional probability mass functions $p_{X_s|\mathbf{Y}}(k|\mathbf{y}, \boldsymbol{\theta}(p))$ are very difficult to obtain when $\mathbf{X}$ is a MRF; in fact, these are the quantities we are trying to approximate to obtain $\mathbf{x}_{MPM}$. In general, this difficulty in obtaining $Q(\boldsymbol{\theta}, \boldsymbol{\theta}(p-1))$ is encountered when the hidden data are modeled as a MRF.

4 EM/MPM ALGORITHM

The EM/MPM algorithm proposed in this paper combines the techniques described in Sections 3.1 and 3.2. First, the MPM algorithm is performed using an initial estimate of $\boldsymbol{\theta}$, say $\hat{\boldsymbol{\theta}}(0)$. After a certain number of iterations of the MPM algorithm, the resulting estimates of $p_{X_s|\mathbf{Y}}(k|\mathbf{y}, \hat{\boldsymbol{\theta}}(0))$ are used in Equations 12 through 14 to obtain an updated estimate of $\boldsymbol{\theta}$, say $\hat{\boldsymbol{\theta}}(1)$. This new estimate of $\boldsymbol{\theta}$ is then used in the MPM algorithm to find estimates of $p_{X_s|\mathbf{Y}}(k|\mathbf{y}, \hat{\boldsymbol{\theta}}(1))$, which are then used to update the estimate of $\boldsymbol{\theta}$. This process is continued until some suitable stopping point is reached.

The EM/MPM algorithm generates a (finite) collection of Markov chains $\mathbf{X}(1,t), \mathbf{X}(2,t), \ldots, \mathbf{X}(P+1,t)$, for some $P \geq 1$. Generation of $\mathbf{X}(p,t)$ is referred to as stage p of the algorithm. The estimate of $\boldsymbol{\theta}$ obtained during stage p is denoted by the random variable $\boldsymbol{\Theta}(p)$. The algorithm begins with the estimate $\boldsymbol{\Theta}(0) = \hat{\boldsymbol{\theta}}(0)$ for some $\hat{\boldsymbol{\theta}}(0) \in \Omega_\theta$. The Markov chain $\mathbf{X}(1,t)$ is generated using the procedure described in Section 3.1. The state of $X_{q_t}(1,t)$ is determined by sampling from the conditional probability mass function $p_{X_{q_t}|\mathbf{Y}, X_r, r \in \mathcal{G}_{q_t}, \boldsymbol{\Theta}(0)}(k|\mathbf{y}, x_r(1,t-1), r \in \mathcal{G}_{q_t}, \hat{\boldsymbol{\theta}}(0))$, where $p_{\mathbf{X}|\mathbf{Y}, \boldsymbol{\Theta}(p)}(\mathbf{x}|\mathbf{y}, \hat{\boldsymbol{\theta}}(p))$, for any $p \geq 1$, has the same form as $p_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta})$ given by Equation 5, with

$\boldsymbol{\Theta}(p)$ random, unlike $\boldsymbol{\theta}$, which is deterministic. Thus, if $\hat{\boldsymbol{\theta}}(p) = [\hat{\mu}_1(p), \hat{\sigma}_1^2(p), \ldots, \hat{\mu}_L(p), \hat{\sigma}_L^2(p)]$, then

$$p_{\mathbf{X}|\mathbf{Y},\boldsymbol{\Theta}(p)}(\mathbf{x}|\mathbf{y}, \hat{\boldsymbol{\theta}}(p)) = \frac{1}{z f_{\mathbf{Y},\boldsymbol{\Theta}(p)}(\mathbf{y}|\hat{\boldsymbol{\theta}}(p))} \left[\prod_{r=1}^{N} \frac{1}{\sqrt{2\pi\hat{\sigma}_{x_r}^2(p)}}\right] \cdot$$

$$\exp\left(-\sum_{r=1}^{N} \frac{(y_r - \hat{\mu}_{x_r}(p))^2}{2\hat{\sigma}_{x_r}^2(p)} - \sum_{\{r,s\}\in\mathcal{C}} \beta t(x_r, x_s) - \sum_{\{r\}\in\mathcal{C}} \gamma_{x_r}\right) \tag{15}$$

After each pixel has been visited T_1 times, for some $T_1 \geq 1$, the estimate $\boldsymbol{\Theta}(1)$ is computed. Using the estimates $v_{k,s}(1,t)$ defined by

$$v_{k,s}(1,t) = \frac{1}{t}\sum_{i=1}^{t} u_{k,s}(1,i) \tag{16}$$

where

$$u_{k,s}(1,t) = \begin{cases} 1 & \text{if } X_s(1,t) = k \\ 0 & \text{if } X_s(1,t) \neq k \end{cases} \tag{17}$$

the estimate $\boldsymbol{\Theta}(1) = [M_1(1), S_1(1), \ldots, M_L(1), S_L(1)]$ is computed using

$$M_k(1) = \frac{\sum_{s=1}^{N} y_s v_{k,s}(1,T_1)}{\sum_{s=1}^{N} v_{k,s}(1,T_1)} \tag{18}$$

and

$$S_k(1) = \frac{\sum_{s=1}^{N} (y_s - M_k(1))^2 v_{k,s}(1,T_1)}{\sum_{s=1}^{N} v_{k,s}(1,T_1)} \tag{19}$$

Note that these equations have the same form as the EM update equations (Equations 12 and 13), with $p_{X_s|\mathbf{Y}}(k|\mathbf{y},\boldsymbol{\theta})$ replaced by the estimates $v_{k,s}(1,T_1)$.

In general, the Markov chain $\mathbf{X}(p,t)$ is generated using the Gibbs sampler and sampling from

the distribution $p_{\mathbf{X}|\mathbf{Y},\boldsymbol{\Theta}(p-1)}(\mathbf{x}|\mathbf{y},\hat{\boldsymbol{\theta}}(p-1))$, and the estimate $\boldsymbol{\Theta}(p)$ is computed using the equations

$$M_k(p) = \frac{\sum_{s=1}^{N} y_s v_{k,s}(p, T_p)}{\sum_{s=1}^{N} v_{k,s}(p, T_p)} \tag{20}$$

and

$$S_k(p) = \frac{\sum_{s=1}^{N} (y_s - M_k(p))^2 v_{k,s}(p, T_p)}{\sum_{s=1}^{N} v_{k,s}(p, T_p)} \tag{21}$$

The final estimate of $\boldsymbol{\theta}$ is $\boldsymbol{\Theta}(P)$. The final segmentation is obtained by maximizing over all k the value $v_{k,s}(P+1, T_{P+1})$ for every $s \in S$, for some $T_{P+1} \geq 1$.

The algorithm can be summarized as follows: First, initial estimates $\hat{\boldsymbol{\theta}}(0)$ and $\mathbf{X}(1,0)$ of $\boldsymbol{\theta}$ and $\mathbf{X}$, respectively, are selected. Then, for $p = 1, \ldots, P$, stage p of the algorithm consists of two steps:

1. Perform T_p iterations of the MPM algorithm using $\hat{\boldsymbol{\theta}}(p-1)$ as the value of $\boldsymbol{\theta}$.

2. Use the EM update equations for $\boldsymbol{\theta}$ to obtain $\hat{\boldsymbol{\theta}}(p)$, using the values $v_{k,s}(p, T_p)$ as estimates of $p_{X_s|Y}(k|\mathbf{y}, \boldsymbol{\theta}(p-1))$.

After $\hat{\boldsymbol{\theta}}(P)$ has been obtained as the final estimate of $\boldsymbol{\theta}$, T_{P+1} iterations of the MPM algorithm are performed, using $\hat{\boldsymbol{\theta}}(P)$. The final segmentation is $\mathbf{X}(P+1, T_{P+1})$.

5 CONVERGENCE ANALYSIS

In this section we present analytical results relative to the EM/MPM algorithm. We show two important results. First, we show that the estimates of the marginal probabilities of the class labels obtained during a given stage of the EM/MPM procedure converge with probability 1 to the true values of the class label probabilities, given the estimates of the model parameters obtained during the previous stage. Second, we show that the parameter estimates resulting from the EM/MPM procedure can be made arbitrarily close to the EM estimates of the parameters with probability 1, if a sufficient number of iterations is performed. These two results are significant because they imply that the final estimates of the marginal probabilities of the class labels obtained using the

EM/MPM procedure will be close to the true values of the marginal probabilities of the class labels, to the extent that the EM estimates of the model parameters are close to the true values of the model parameters. This is important because these estimates are maximized to obtain the segmented image, and if the estimates are not close to the true values, then the segmentation algorithm cannot be expected to perform well.

The analysis is presented for a more general case than the formulation of Section 4. For the purpose of analysis $\mathbf{X}$ is a collection of random variables which have a Gibbs distribution, and the EM equations form a sequence $\{\boldsymbol{\theta}(p), p \geq 1\}$, where, for each p, $\boldsymbol{\theta}(p)$ is a function of $p_{X_s|\mathbf{Y}}(k|\mathbf{y}, \boldsymbol{\theta}(p-1))$. The general EM/MPM procedure studied in this section consists of the following steps at stage p:

1. Generate a Markov chain $\mathbf{X}(p,t)$ with limiting distribution $p_{\mathbf{X}|\mathbf{Y},\boldsymbol{\Theta}(p-1)}(\mathbf{x}|\mathbf{y}, \hat{\boldsymbol{\theta}}(p-1))$, where $\boldsymbol{\Theta}(p-1)$ is the estimate of $\boldsymbol{\theta}$ obtained in the previous stage.

2. Approximate the class label probabilities using the equations

$$p_{X_s|\mathbf{Y}}(k|\mathbf{y}, \boldsymbol{\theta}) \approx \frac{1}{T_p} \sum_{t=1}^{T_p} u_{k,s}(p,t) \tag{22}$$

   where $u_{k,s}(p,t)$ is 1 if $X_s(p,t) = k$ and 0 otherwise.

3. Use the EM update equations to compute $\boldsymbol{\Theta}(p)$, the estimate of $\boldsymbol{\theta}$ obtained during stage p, using the class label probability estimates from Step 2 in the EM equations.

After stage P has been completed, the final estimate of $\boldsymbol{\theta}$, $\boldsymbol{\Theta}(P)$, is computed using the EM update equations, and the MPM algorithm is performed once more, using the final estimate of $\boldsymbol{\theta}$, to obtain the final estimates of the marginal class label probabilities, and then the final segmentation.

We first describe convergence of the MPM algorithm when the value of $\boldsymbol{\theta}$ is known. The analysis of the EM/MPM procedure is then presented.

5.1 Convergence of the MPM Algorithm

To determine the convergence properties of the right-hand side of Equation 9 when $\boldsymbol{\theta}$ is known, we use Theorem C from [10], which states that if there exists a τ such that $S \subseteq \{q_{t+1}, \ldots, q_{t+\tau}\}$ for

all t, then for any function $\mathbf{g}$ on $\Omega_{\mathbf{X}}$ and for any starting configuration $\mathbf{x}(0) \in \Omega_{\mathbf{X}}$,

$$P(\lim_{t\to\infty} \frac{1}{t}\sum_{i=1}^{t} \mathbf{g}(\mathbf{X}(i)) = \int_{\Omega_{\mathbf{X}}} \mathbf{g}(\mathbf{x})\, d\pi(\mathbf{x}) | \mathbf{X}(0) = \mathbf{x}(0)) = 1 \tag{23}$$

where π is the limiting distribution of $\mathbf{X}(t)$ (in our case, $\pi(\mathbf{x}) = p_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta})$). The condition $S \subseteq \{q_{t+1}, \ldots, q_{t+\tau}\}$ is easily satisfied. For example, if the pixels are visited in raster-scan order, then $\tau = N$. Let $\mathbf{g}(\mathbf{x})$ be a vector-valued function with elements

$$g_{k,s}(\mathbf{x}) = \begin{cases} 1 & \text{if } x_s = k \\ 0 & \text{if } x_s \neq k \end{cases} \tag{24}$$

ordered by a lexicographical ordering of k, s to form a vector. Then

$$\frac{1}{t}\sum_{i=1}^{t} g_{k,s}(\mathbf{X}(i)) = \frac{1}{t}\sum_{i=1}^{t} u_{k,s}(i) \tag{25}$$

and

$$\int_{\Omega_{\mathbf{X}}} g_{k,s}(\mathbf{x})\, d\pi(\mathbf{x}) = \sum_{\mathbf{x}\in\Omega_{k,s}} p_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta}) = p_{X_s|\mathbf{Y}}(k|\mathbf{y}, \boldsymbol{\theta}) \tag{26}$$

Hence, using Equation 23,

$$P(\lim_{t\to\infty} \frac{1}{t}\sum_{i=1}^{t} u_{k,s}(i) = p_{X_s|\mathbf{Y}}(k|\mathbf{y}, \boldsymbol{\theta})\ \forall k, s | \mathbf{X}(0) = \mathbf{x}(0)) = 1 \tag{27}$$

Thus, for any initial configuration, the estimates $\frac{1}{t}\sum_{i=1}^{t} u_{k,s}(i)$ converge to $p_{X_s|\mathbf{Y}}(k|\mathbf{y}, \boldsymbol{\theta})$ for all k, s, with probability 1.

5.2 Analysis of the EM/MPM Procedure

We present two theorems in this section. The first theorem describes convergence of the class label probability estimates; the second theorem relates to the parameter estimates. The proofs of these theorems are provided in Appendix A and Appendix B.

5.2.1 Convergence of the class label probability estimates

Theorem 1 *For every* $p = 1, \ldots, P+1$, *and* $\forall k, s$,

$$\frac{1}{t}\sum_{i=1}^{t} u_{k,s}(p,i) \longrightarrow p_{X_s|\mathbf{Y},\boldsymbol{\Theta}_{(p-1)}}(k|\mathbf{y}, \hat{\boldsymbol{\theta}}(p-1)) \tag{28}$$

as $t \longrightarrow \infty$ *with probability 1, given that* $\boldsymbol{\Theta}(p-1) = \hat{\boldsymbol{\theta}}(p-1)$.

Stages $p = 1, \ldots, P$ of the EM/MPM procedure are used to obtain the final parameter estimate $\boldsymbol{\Theta}(P)$, and stage $P+1$ is used to obtain the segmented image. The segmented image is obtained using the estimate $\boldsymbol{\Theta}(P)$ for $\boldsymbol{\theta}$ and the final estimates of the class label probabilities given by the left-hand side of Equation 28 with $p = P+1$. Thus, Theorem 1 implies that the final estimates of the marginal probabilities of the class labels obtained using the EM/MPM procedure will be close to the true values of the marginal probabilities of the class labels evaluated with the model parameters equal to the final estimate of $\boldsymbol{\theta}$, conditioned on the final estimate of $\boldsymbol{\theta}$.

5.2.2 Analysis of the parameter estimates

Theorem 2 *If* $\boldsymbol{\theta}(p)$ *is a continuous function of* $p_{X_s|\mathbf{Y}}(k|\mathbf{y}, \boldsymbol{\theta}(p-1))$ *and* $p_{X_s|\mathbf{Y}}(k|\mathbf{y}, \boldsymbol{\theta})$ *is a continuous function of* $\boldsymbol{\theta}$ *for every* k, s, *and if* $\boldsymbol{\theta}^*$ *is the EM estimate of* $\boldsymbol{\theta}$, *then for any* $\varepsilon > 0$, $|\boldsymbol{\Theta}(P) - \boldsymbol{\theta}^*| < \varepsilon$ *if* P *and* $T_p, 1 \leq p \leq P$ *are chosen sufficiently large.*

It can be shown that the continuity conditions required for Theorem 2 to hold are satisfied for the image models used in this paper. Theorem 2 implies that the parameter estimates resulting from the EM/MPM procedure can be made arbitrarily close to the EM estimates of the parameters with probability 1, if a sufficient number of iterations is performed. It should be noted that there is no guarantee that the number of iterations required at each stage for this result to hold are finite, but the analysis provides theoretical evidence that a sufficient number of iterations of the Gibbs sampler at each stage will provide an estimate of $\boldsymbol{\theta}$ close to the EM estimate. Theorems 1 and 2 together imply that the final estimates of the marginal probabilities of the class labels obtained using the EM/MPM procedure will be close to the true values of the marginal probabilities of the class labels, to the extent that the EM estimate of $\boldsymbol{\theta}$ is close to the true value of $\boldsymbol{\theta}$.

6 EXPERIMENTAL RESULTS

The EM/MPM algorithm was applied to several different types of imagery. Results demonstrating the performance of the algorithm on synthetic imagery, infrared imagery, and mammographic imagery are presented in this section. For all results presented in this section, the value of the spatial interaction parameter β was assumed to be 2.4, and T_p, the number of iterations of the Gibbs sampler at stage p, was set to 3 for every $p = 1, \ldots, P+1$. Unless otherwise noted, the relative cost of class k (i.e., γ_k in Equation 1) is assumed to be zero, for every k. Also, unless otherwise noted, initial estimates of the class means and variances were obtained using

$$\mu_k(0) = \frac{128}{L} + \frac{255k}{L} \tag{29}$$

for the means and setting the variance for each class to 20. For every pixel $s \in S$ the initial estimate of X_s was chosen to be uniformly distributed over all classes, and independent of the initial estimates at other pixels.

Figure 1(a) shows the first test image. This image is a composite of Brodatz textures wood and grass. The segmented image obtained from 70 stages of the EM/MPM algorithm is shown in Figure 1(b), and the segmentation after 500 iterations of the deterministic EM/MPM algorithm proposed in [6] is shown in Figure 1(c). It can be seen that for this image the EM/MPM algorithm provides better performance than the deterministic EM/MPM algorithm.

The second test image, shown in Figure 2(a), is a composite of Brodatz textures cork and leather. The segmentation obtained after 300 stages of the EM/MPM algorithm is shown in Figure 2(b), and the segmentation after 500 iterations of the deterministic EM/MPM algorithm is shown in Figure 2(c). For this image the EM/MPM algorithm performs significantly better than the deterministic EM/MPM algorithm. The EM/MPM algorithm does have difficulty correctly classifying the pixels in the cork region. This is because the observed image model does not fit the cork texture well. We have developed a multiresolution extension of the EM/MPM algorithm to address this issue [15].

Table 1 shows the percentage of pixels which were misclassified by the EM/MPM and deterministic EM/MPM algorithms for the two synthetic test images. For both images the EM/MPM algorithm performed better than the deterministic EM/MPM algorithm in terms of minimizing the

Table 1: Percentage of Misclassified Pixels

Image	EM/MPM	Deterministic EM/MPM
Figure 1(a)	2.4	5.0
Figure 2(a)	4.8	40.0

Table 2: Parameter Estimates for Synthetic Test Image

	μ_1	σ_1	μ_2	σ_2
Sample Mean/Sample Standard Deviation	83.0	21.6	122.7	54.6
Initial Estimate	64.0	4.5	191.0	4.5
Final EM/MPM Estimate	81.7	20.7	122.9	53.6
Final Deterministic EM/MPM Estimate	80.9	21.5	126.6	52.8

number of misclassified pixels.

Figures 3 and 4 illustrate intermediate results as the EM/MPM and deterministic EM/MPM algorithms segment the image shown in Figure 1(a). Figures 3(a), (b), (c), and (d) show results after 30, 50, 70, and 300 stages, respectively, of the EM/MPM algorithm, and Figures 4(a), (b), (c), and (d) show results after 10, 20, 500, and 800 iterations, respectively, of the deterministic EM/MPM algorithm [6]. Table 2 shows the final parameter estimates obtained after 300 stages of the EM/MPM algorithm and 800 iterations of the deterministic EM/MPM algorithm, as well as the sample means and sample standard deviations computed using the true segmentation. The EM/MPM estimates of the means for the two classes are closer to the sample means than the deterministic EM/MPM estimates. The accuracy of the estimates of the means is critical to the performance of both algorithms.

The algorithm was also tested on infrared imagery. Figures 5 through 7 show three infrared images which were segmented using the EM/MPM algorithm, and the resulting segmentations. In each of these figures the image shown in (b) is the segmentation resulting from 70 stages of EM/MPM with $L = 2$ and the image shown in (c) is the result after 70 stages with $L = 3$. The segmentations obtained using the deterministic EM/MPM algorithm on the infrared images were similar to the segmentations obtained using the EM/MPM algorithm except for the image shown in Figure 7. The result after 500 iterations of the deterministic EM/MPM algorithm for this image

is shown in Figure 8(b). For this image the deterministic EM/MPM algorithm does not perform as well as the EM/MPM algorithm.

To segment mammography images, we use the *a priori* knowledge that a tumor is expected to cover a relatively small region compared to the normal tissue and the background. This information is incorporated into the label field model by using nonzero values for the parameters γ_k. The effect of this is to increase the cost at each pixel of belonging to the class corresponding to the tumor, as described in Section 2.1. Also, the fact that a tumor is expected to be relatively small and the fact that a tumor is usually associated with higher grayscale values than the other regions are used to compute initial estimates of the class means and variances. The grayscale values in the observed image are sorted, and the sample mean and variance of the highest grayscale values are used for the initial parameter estimates for the tumor class. The remaining values are used to compute initial estimates of the parameters for the normal tissue and background region.

We assume that mammography images consist of three classes: background, normal tissue, and tumor. The values used for the class cost parameters were $\gamma_1 = \gamma_2 = 2.3$ and $\gamma_3 = 5$, where class 3 is the tumor class. These values were determined experimentally using a variety of sample mammography images.

The mammography image which was tested is shown in Figure 9(a). The corresponding truth image is shown in Figure 9(b). The segmented image obtained after 100 stages of the EM/MPM algorithm is shown in Figure 9(c). It can be seen that the algorithm segmented the three regions quite well.

The final issue that we discuss is a comparison of the amount of computation required to achieve convergence for two of the test images using the EM/MPM and deterministic EM/MPM algorithms. We consider the convergence of the parameter estimates to do this, although the amount of computation could also be measured by studying convergence of the MPM algorithm at each stage of the EM/MPM algorithm.

We use the number of visits per pixel as a measure of computational complexity because the number of operations required at each pixel visit are comparable for the two algorithms under consideration. Table 3 shows the number of visits to each pixel required for convergence of the parameter estimates obtained by the two algorithms for the two test images in Figures 1(a) and 2(a). It can be seen that the deterministic EM/MPM algorithm converges more quickly than

Table 3: Visits Per Pixel

Image	EM/MPM	Deterministic EM/MPM
Figure 1(a)	1250	2160
Figure 2(a)	3000	1050

the EM/MPM algorithm for the second image, but for the first image the EM/MPM algorithm converges more quickly. Although the deterministic EM/MPM algorithm converged more quickly for the experimental results presented in [6], we have found that this is not necessarily the case for all images. This result is not completely surprising, because the EM algorithm is known to be slow to converge in some cases, and, unlike the EM/MPM algorithm, the deterministic algorithm is not guaranteed to converge to the correct values of the class label probabilities which are used to compute the EM updates.

7 Conclusion

The analysis of the EM/MPM algorithm described in this paper implies that the algorithm can be expected to minimize the expected value of the number of misclassified pixels, to the extent that the EM estimates of the model parameters are close to the true values of the model parameters.

We presented experimental results demonstrating the performance of the algorithm, and compared these results with those obtained by the deterministic EM/MPM algorithm. Our results for synthetic images showed that the EM/MPM algorithm performs better than the deterministic EM/MPM algorithm in terms of minimizing the number of misclassified pixels. Also, the deterministic EM/MPM algorithm does not necessarily converge more quickly than the EM/MPM algorithm.

A postscript version of this paper is available via anonymous ftp to `skynet.ecn.purdue.edu` (Internet address `128.46.154.48`) in the directory `/pub/dist/delp/segment`.

APPENDIX: Appendix A

Theorem 1 *For every* $p = 1, \ldots, P+1$, *and* $\forall k, s$,

$$\frac{1}{t}\sum_{i=1}^{t} u_{k,s}(p,i) \longrightarrow p_{X_s|\mathbf{Y},\boldsymbol{\Theta}_{(p-1)}}(k|\mathbf{y},\hat{\boldsymbol{\theta}}(p-1)) \tag{30}$$

as $t \longrightarrow \infty$ *with probability 1, given that* $\boldsymbol{\Theta}(p-1) = \hat{\boldsymbol{\theta}}(p-1)$.

Proof: We know that, by construction,

$$P(\mathbf{X}(1,t) = \mathbf{x}(1,t)|\mathbf{X}(1,t-1) = \mathbf{x}(1,t-1), \mathbf{Y} = \mathbf{y}, \boldsymbol{\Theta}(0) = \hat{\boldsymbol{\theta}}(0)) =$$

$$p_{X_{q_t}|X_s, s\neq q_t, \mathbf{Y}, \boldsymbol{\Theta}_{(0)}}(x_{q_t}(1,t)|x_s(1,t), s \neq q_t, \mathbf{y}, \hat{\boldsymbol{\theta}}(0)) \tag{31}$$

if $x_s(1,t) = x_s(1,t-1)\ \forall s \neq q_t$. If $x_s(1,t) \neq x_s(1,t-1)$ for any $s \neq q_t$ then the probability on the left-hand side of Equation 31 is equal to 0. Following the proof of Theorem A in [10], it can be shown that, if $\{q_t, t \geq 1\}$ contains every $s \in S$ infinitely often then for any $\mathbf{x}(1,0) \in \Omega_{\mathbf{X}}$, $\mathbf{y} \in \Omega_{\mathbf{Y}}$, $\hat{\boldsymbol{\theta}}(0) \in \Omega_\theta$, and $\mathbf{x} \in \Omega_{\mathbf{X}}$,

$$\lim_{t\to\infty} P(\mathbf{X}(1,t) = \mathbf{x}|\mathbf{X}(1,0) = \mathbf{x}(1,0), \mathbf{Y} = \mathbf{y}, \boldsymbol{\Theta}(0) = \hat{\boldsymbol{\theta}}(0)) = p_{\mathbf{X}|\mathbf{Y},\boldsymbol{\Theta}_{(0)}}(\mathbf{x}|\mathbf{y},\hat{\boldsymbol{\theta}}(0)) \tag{32}$$

Also, using Theorem C from [10] as described in Section 5.1,

$$P(\lim_{t\to\infty} v_{k,s}(1,t) = p_{X_s|\mathbf{Y},\boldsymbol{\Theta}_{(0)}}(k|\mathbf{y},\hat{\boldsymbol{\theta}}(0))|\mathbf{X}(1,0) = \mathbf{x}(1,0), \mathbf{Y} = \mathbf{y}, \boldsymbol{\Theta}(0) = \hat{\boldsymbol{\theta}}(0) \forall k, s)$$

$$= 1 \tag{33}$$

for any $\mathbf{x}(1,0) \in \Omega_{\mathbf{X}}, \mathbf{y} \in \Omega_{\mathbf{Y}}, \hat{\boldsymbol{\theta}}(0) \in \Omega_\theta$.

An important consequence of the method used to construct $\mathbf{X}(1,t), \mathbf{X}(2,t), \ldots, \mathbf{X}(P+1,t)$ is that for any $p \in \{2, \ldots, P+1\}$ and any $m_p \geq 1$, the random variables in $\mathbf{X}(p,1), \mathbf{X}(p,2), \ldots, \mathbf{X}(p,m_p)$ are conditionally independent of the random variables in stages $1, 2, \ldots, p-1$, given $\boldsymbol{\Theta}(p-1)$. This means that, if A_p denotes the event $\{\mathbf{X}(p\text{–}1,0) = \mathbf{x}(p\text{–}1,0), \ldots, \mathbf{X}(p\text{–}1,m_{p\text{–}1}) = \mathbf{x}(p\text{–}1,m_{p\text{–}1}), \boldsymbol{\Theta}(p\text{–}2) = \hat{\boldsymbol{\theta}}(p\text{–}2), \mathbf{X}(p\text{–}2,0) = \mathbf{x}(p\text{–}2,0), \ldots, \mathbf{X}(p\text{–}2,m_{p\text{–}2}) = \mathbf{x}(p\text{–}2,m_{p\text{–}2}), \ldots, \boldsymbol{\Theta}(1) = \hat{\boldsymbol{\theta}}(1), \mathbf{X}(1,0) =$

$\mathbf{x}(1,0), \ldots, \mathbf{X}(1,m_1) = \mathbf{x}(1,m_1), \boldsymbol{\Theta}(0) = \boldsymbol{\theta}(0)\}$ and B_p denotes the event $\{\boldsymbol{\Theta}(p-1) = \hat{\boldsymbol{\theta}}(p-1)\}$, then

$$P(\mathbf{X}(p,1) = \mathbf{x}(p,1), \ldots, \mathbf{X}(p,m_p) = \mathbf{x}(p,m_p) | \mathbf{X}(p,0) = \mathbf{x}(p,0), \mathbf{Y} = \mathbf{y}, A_p, B_p) =$$

$$P(\mathbf{X}(p,1) = \mathbf{x}(p,1), \ldots, \mathbf{X}(p,m_p) = \mathbf{x}(p,m_p) | \mathbf{X}(p,0) = \mathbf{x}(p,0), \mathbf{Y} = \mathbf{y}, B_p) \tag{34}$$

Now, by construction, for any p,

$$P(\mathbf{X}(p,t) = \mathbf{x}(p,t) | \mathbf{X}(p,t-1) = \mathbf{x}(p,t-1), \mathbf{Y} = \mathbf{y}, B_p) =$$

$$p_{X_{q_t} | X_s, s \neq q_t, \mathbf{Y}, \boldsymbol{\Theta}(p-1)}(x_{q_t}(p,t) | x_s(p,t), \mathbf{y}, \hat{\boldsymbol{\theta}}(p-1)) \tag{35}$$

if $x_s(p,t) = x_s(p,t-1)\ \forall s \neq q_t$. If $x_s(p,t) \neq x_s(p,t-1)$ for any $s \neq q_t$ then the probability on the left-hand side of Equation 35 is equal to 0. Hence, using Theorems A and C from [10],

$$P(\lim_{t \to \infty} v_{k,s}(p,t) = p_{X_s | \mathbf{Y}, \boldsymbol{\Theta}(p-1)}(k | \mathbf{y}, \hat{\boldsymbol{\theta}}(p-1)) \forall k, s | \mathbf{X}(p,0) = \mathbf{x}(p,0), \mathbf{Y} = \mathbf{y}, B_p)$$

$$= 1 \tag{36}$$

for any $\mathbf{x}(p,0) \in \Omega_{\mathbf{X}}, \mathbf{y} \in \Omega_{\mathbf{y}}$, and $\hat{\boldsymbol{\theta}}(p-1) \in \Omega_\theta$ and, by virtue of Equation 34, the estimates $v_{k,s}(p,t)$ converge with probability 1 to $p_{X_s | \mathbf{Y}, \boldsymbol{\Theta}(p-1)}(k | \mathbf{y}, \hat{\boldsymbol{\theta}}(p-1))$, given $\boldsymbol{\Theta}(p-1) = \hat{\boldsymbol{\theta}}(p-1)$, independently of all other random variables in stages $1, \ldots, p-1$. This completes the proof.

APPENDIX: Appendix B

Theorem 2 *If $\boldsymbol{\theta}(p)$ is a continuous function of $p_{X_s | \mathbf{Y}}(k | \mathbf{y}, \boldsymbol{\theta}(p-1))$ and $p_{X_s | \mathbf{Y}}(k | \mathbf{y}, \boldsymbol{\theta})$ is a continuous function of $\boldsymbol{\theta}$ for every k, s, and if $\boldsymbol{\theta}^*$ is the EM estimate of $\boldsymbol{\theta}$, then for any $\varepsilon > 0$, $|\boldsymbol{\Theta}(P) - \boldsymbol{\theta}^*| < \varepsilon$ if P and $T_p, 1 \leq p \leq P$ are chosen sufficiently large.*

Proof. If $\{\boldsymbol{\theta}(p), p \geq 1\}$ is the sequence of estimates obtained using the EM algorithm when the values of $p_{X_s | \mathbf{Y}}(k | \mathbf{y}, \boldsymbol{\theta})$ are known, suppose that $\boldsymbol{\theta}(p)$ converges to $\boldsymbol{\theta}^*$ for some $\boldsymbol{\theta}^* \in \Omega_\theta$. We would like for $\boldsymbol{\Theta}(P)$ to be close to $\boldsymbol{\theta}^*$ in some sense. In particular, for an arbitrarily small but positive ε we address the question of whether $|\boldsymbol{\Theta}(P) - \boldsymbol{\theta}^*|$ can be made less than ε by making P and $T_1, \ldots, T_P$

large enough.

We define the notation $a_{k,s}(\boldsymbol{\theta}) = p_{X_s|\mathbf{Y}}(k|\mathbf{y}, \boldsymbol{\theta})$. (Note that the dependence on $\mathbf{y}$ is suppressed in this notation). We assume that $\boldsymbol{\theta}(p)$ is a continuous function of $a_{k,s}(\boldsymbol{\theta}(p-1))$ for every k, s and that $a_{k,s}(\boldsymbol{\theta})$ is a continuous function of $\boldsymbol{\theta}$, for every k, s. We also assume that $\boldsymbol{\Theta}(0) = \boldsymbol{\theta}(0)$, i.e., the EM/MPM procedure is started with the same initial estimate of $\boldsymbol{\theta}$ as the EM sequence that converges to $\boldsymbol{\theta}^*$.

Fix $\varepsilon > 0$. Since $\boldsymbol{\theta}(p) \longrightarrow \boldsymbol{\theta}^*$, there exists a $P >= 1$ such that $|\boldsymbol{\theta}(p) - \boldsymbol{\theta}^*| < \varepsilon/2$ for all $p \geq P$. Fix P to be such a value. Then, to make $|\boldsymbol{\Theta}(P) - \boldsymbol{\theta}^*| < \varepsilon$, it is sufficient to make $|\boldsymbol{\Theta}(P) - \boldsymbol{\theta}(P)| < \varepsilon/2$, since

$$|\boldsymbol{\Theta}(P) - \boldsymbol{\theta}^*| \leq |\boldsymbol{\Theta}(P) - \boldsymbol{\theta}(P)| + |\boldsymbol{\theta}(P) - \boldsymbol{\theta}^*| \tag{37}$$

and, hence,

$$|\boldsymbol{\Theta}(P) - \boldsymbol{\theta}(P)| < \frac{\varepsilon}{2} \Longrightarrow |\boldsymbol{\Theta}(P) - \boldsymbol{\theta}^*| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \tag{38}$$

Since $\boldsymbol{\theta}(p)$ is a continuous function of $a_{k,s}(\boldsymbol{\theta}(p-1))$, there exists a $\delta_P > 0$ such that

$$|v_{k,s}(P, T_P) - a_{k,s}(\boldsymbol{\theta}(P-1))| < \delta_P \ \forall k, s \Longrightarrow |\boldsymbol{\Theta}(P) - \boldsymbol{\theta}(P)| < \frac{\varepsilon}{2} \tag{39}$$

Fix δ_P to satisfy this condition. For any $\hat{\boldsymbol{\theta}}(P-1) \in \Omega_\theta$, by the triangle inequality,

$$|v_{k,s}(P, T_P) - a_{k,s}(\boldsymbol{\theta}(P-1))| \leq$$

$$|v_{k,s}(P, T_P) - a_{k,s}(\hat{\boldsymbol{\theta}}(P-1))| + |a_{k,s}(\hat{\boldsymbol{\theta}}(P-1)) - a_{k,s}(\boldsymbol{\theta}(P-1))| \tag{40}$$

Looking at the first term on the right-hand side of Equation 40, we recall that $v_{k,s}(P, t)$ converges to $a_{k,s}(\hat{\boldsymbol{\theta}}(P-1))$ with probability 1, given that $\boldsymbol{\Theta}(P-1) = \hat{\boldsymbol{\theta}}(P-1)$. Considering the second term on the right-hand side of Equation 40, since, for every k, s, $a_{k,s}(\boldsymbol{\theta})$ is continuous at $\boldsymbol{\theta} = \boldsymbol{\theta}(P)$, there exists an ε_{P-1} such that

$$|\boldsymbol{\Theta}(P-1) - \boldsymbol{\theta}(P-1)| < \varepsilon_{P-1} \Longrightarrow |a_{k,s}(\boldsymbol{\Theta}(P-1)) - a_{k,s}(\boldsymbol{\theta}(P-1))| < \frac{\delta_P}{2} \ \forall k, s \tag{41}$$

Fix ε_{P-1} to be such a value. We choose a set of values $\{\varepsilon_p, \delta_p, 1 \leq p \leq P\}$, where $\varepsilon_P = \varepsilon/2$, and

δ_P and ε_{P-1} are fixed as described above. For $p = P-1, P-2, \ldots, 2$, $\delta_p > 0$ is chosen to satisfy

$$|v_{k,s}(p, T_p) - a_{k,s}(\boldsymbol{\theta}(p-1))| < \delta_p \; \forall k, s \Longrightarrow |\boldsymbol{\Theta}(p) - \boldsymbol{\theta}(p)| < \varepsilon_p \tag{42}$$

and $\varepsilon_{p-1} > 0$ is chosen to satisfy

$$|\boldsymbol{\Theta}(p-1) - \boldsymbol{\theta}(p-1)| < \varepsilon_{p-1} \Longrightarrow |a_{k,s}(\boldsymbol{\Theta}(p-1)) - a_{k,s}(\boldsymbol{\theta}(p-1))| < \frac{\delta_p}{2} \tag{43}$$

Finally, the value δ_1 is chosen such that

$$|v_{k,s}(1, T_1) - a_{k,s}(\boldsymbol{\theta}(0))| < \delta_1 \; \forall k, s \Longrightarrow |\boldsymbol{\Theta}(1) - \boldsymbol{\theta}(1)| < \varepsilon_1 \tag{44}$$

From Equation 44, it can be seen that

$$P(|\boldsymbol{\Theta}(1) - \boldsymbol{\theta}(1)| < \varepsilon_1 | \boldsymbol{\Theta}(0) = \boldsymbol{\theta}(0)) \geq P(|v_{k,s}(1, T_1) - a_{k,s}(\boldsymbol{\theta}(0))| < \delta_1 \forall k, s | \boldsymbol{\Theta}(0) = \boldsymbol{\theta}(0)) \tag{45}$$

Recall that, for every k, s, $v_{k,s}(1, t)$ converges to $a_{k,s}(\boldsymbol{\theta}(0))$ with probability 1, given that $\boldsymbol{\Theta}(0) = \boldsymbol{\theta}(0)$. Let $(\Omega, \mathcal{F}, P)$ be the underlying probability space, i.e., Ω is the set of possible outcomes, $\mathcal{F}$ is a σ-algebra in Ω containing all events, and P is a probability measure. Then the event

$$A_{1,\theta(0)} = \{\omega \in \Omega : \lim_{t \to \infty} v_{k,s}(1, t, \omega) = a_{k,s}(\boldsymbol{\theta}(0)) \; \forall k, s\}, \tag{46}$$

where $v_{k,s}(1, t)$ is written as $v_{k,s}(1, t, \omega)$ to explicitly denote its dependence on ω, occurs with probability 1, given that $\boldsymbol{\Theta}(0) = \boldsymbol{\theta}(0)$. Now, for every $\boldsymbol{\theta}(0) \in \Omega_\theta$ and $\omega \in A_{1,\theta(0)}$, there exists a $T(\boldsymbol{\theta}(0), \omega)$ such that, given that $\boldsymbol{\Theta}(0) = \boldsymbol{\theta}(0)$

$$|v_{k,s}(1, t, \omega) - a_{k,s}(\boldsymbol{\theta}(0))| < \delta_1 \; \forall t \geq T(\boldsymbol{\theta}(0), \omega), \forall k, s \tag{47}$$

Letting $T_1 = \sup_{\theta(0) \in \Omega_\theta} \{ \sup_{\omega \in A_{1,\theta(0)}} \{T(\boldsymbol{\theta}(0), \omega)\}\}$ gives

$$A_{1,\theta(0)} \subseteq \{\omega \in \Omega : |v_{k,s}(1, t, \omega) - a_{k,s}(\boldsymbol{\theta}(0))| < \delta_1 \; \forall t \geq T_1, \forall k, s\} \tag{48}$$

for any $\boldsymbol{\theta}(0) \in \Omega_\theta$. Then

$$P(|v_{k,s}(1,T_1) - a_{k,s}(\boldsymbol{\theta}(0))| < \delta_1 \forall k,s | \boldsymbol{\Theta}(0) = \boldsymbol{\theta}(0)) \geq P(A_{1,\theta(0)} | \boldsymbol{\Theta}(0) = \boldsymbol{\theta}(0)) = 1 \tag{49}$$

From Equations 45 and 49, we see that,

$$P(|\boldsymbol{\Theta}(1) - \boldsymbol{\theta}(1)| < \varepsilon_1 | \boldsymbol{\Theta}(0) = \boldsymbol{\theta}(0)) = 1 \ \ \forall \boldsymbol{\theta}(0) \in \Omega_\theta \tag{50}$$

Then

$$P(|\boldsymbol{\Theta}(1) - \boldsymbol{\theta}(1)| < \varepsilon_1) = \int_{\Omega_\theta} P(|\boldsymbol{\Theta}(1) - \boldsymbol{\theta}(1)| < \varepsilon_1 | \boldsymbol{\Theta}(0) = \boldsymbol{\theta}(0)) f_{\boldsymbol{\Theta}(0)}(\boldsymbol{\theta}(0))\, d\boldsymbol{\theta}(0) \tag{51}$$

where $f_{\boldsymbol{\Theta}(0)}(\boldsymbol{\theta}(0))$ is the probability density function of $\boldsymbol{\Theta}(0)$, and from Equation 50,

$$P(|\boldsymbol{\Theta}(1) - \boldsymbol{\theta}(1)| < \varepsilon_1) = \int_{\Omega_\theta} f_{\boldsymbol{\Theta}(0)}(\boldsymbol{\theta}(0))\, d\boldsymbol{\theta}(0) = 1 \tag{52}$$

Next, consider $\boldsymbol{\Theta}(2)$. From Equation 42 with $p = 2$ it can be seen that

$$P(|\boldsymbol{\Theta}(2) - \boldsymbol{\theta}(2)| < \varepsilon_2) \geq P(|v_{k,s}(2,T_2) - a_{k,s}(\boldsymbol{\theta}(1))| < \delta_2 \forall k,s) \tag{53}$$

and, since

$$|v_{k,s}(2,T_2) - a_{k,s}(\boldsymbol{\theta}(1))| \leq |v_{k,s}(2,T_2) - a_{k,s}(\boldsymbol{\Theta}(1))| + |a_{k,s}(\boldsymbol{\Theta}(1)) - a_{k,s}(\boldsymbol{\theta}(1))| \tag{54}$$

we have that

$$\begin{aligned}
&P(|v_{k,s}(2,T_2) - a_{k,s}(\boldsymbol{\theta}(1))| < \delta_2 \ \forall k,s) \geq \\
&P(|v_{k,s}(2,T_2) - a_{k,s}(\boldsymbol{\Theta}(1))| < \frac{\delta_2}{2} \ \forall k,s, \quad |a_{k,s}(\boldsymbol{\Theta}(1)) - a_{k,s}(\boldsymbol{\theta}(1))| < \frac{\delta_2}{2} \ \forall k,s) \\
&= P(|v_{k,s}(2,T_2) - a_{k,s}(\boldsymbol{\Theta}(1))| < \frac{\delta_2}{2} \ \forall k,s \mid |a_{k,s}(\boldsymbol{\Theta}(1)) - a_{k,s}(\boldsymbol{\theta}(1))| < \frac{\delta_2}{2} \ \forall k,s) \ \cdot \\
&P(|a_{k,s}(\boldsymbol{\Theta}(1)) - a_{k,s}(\boldsymbol{\theta}(1))| < \frac{\delta_2}{2} \ \forall k,s)
\end{aligned} \tag{55}$$

From the continuity of $a_{k,s}(\boldsymbol{\theta})$ at $\boldsymbol{\theta} = \boldsymbol{\theta}(1)$,

$$P(|a_{k,s}(\boldsymbol{\Theta}(1)) - a_{k,s}(\boldsymbol{\theta}(1))| < \frac{\delta_2}{2}\ \forall k,s) \geq P(|\boldsymbol{\Theta}(1) - \boldsymbol{\theta}(1))| < \varepsilon_1) = 1 \tag{56}$$

Also,

$$P(|v_{k,s}(2,T_2) - a_{k,s}(\boldsymbol{\Theta}(1))| < \frac{\delta_2}{2}\ \forall k,s \mid |a_{k,s}(\boldsymbol{\Theta}(1)) - a_{k,s}(\boldsymbol{\theta}(1))| < \frac{\delta_2}{2}\ \forall k,s)$$

$$= \int_{\Omega_\theta} P(|v_{k,s}(2,T_2) - a_{k,s}(\boldsymbol{\Theta}(1))| < \frac{\delta_2}{2}\ \forall k,s \mid$$
$$\boldsymbol{\Theta}(1) = \hat{\boldsymbol{\theta}}(1), |a_{k,s}(\boldsymbol{\Theta}(1)) - a_{k,s}(\boldsymbol{\theta}(1))| < \frac{\delta_2}{2}\ \forall k,s)\ \cdot$$

$$f_{\boldsymbol{\Theta}(1)}(\hat{\boldsymbol{\theta}}(1)||a_{k,s}(\boldsymbol{\Theta}(1)) - a_{k,s}(\boldsymbol{\theta}(1))| < \frac{\delta_2}{2}\ \forall k,s))\, d\hat{\boldsymbol{\theta}}(1) \tag{57}$$

The integrand in Equation 57 is zero if $|a_{k,s}(\hat{\boldsymbol{\theta}}(1)) - a_{k,s}(\boldsymbol{\theta}(1))| \geq \frac{\delta_2}{2}$ for any k,s, since the conditional probability density function of $\boldsymbol{\Theta}(1)$ is zero for this case. If $|a_{k,s}(\hat{\boldsymbol{\theta}}(1)) - a_{k,s}(\boldsymbol{\theta}(1))| < \frac{\delta_2}{2}\ \forall k,s$, then

$$P(|v_{k,s}(2,T_2) - a_{k,s}(\boldsymbol{\Theta}(1))| < \frac{\delta_2}{2}\ \forall k,s \mid$$
$$\boldsymbol{\Theta}(1) = \hat{\boldsymbol{\theta}}(1), |a_{k,s}(\boldsymbol{\Theta}(1)) - a_{k,s}(\boldsymbol{\theta}(1))| < \frac{\delta_2}{2}\ \forall k,s) =$$

$$P(|v_{k,s}(2,T_2) - a_{k,s}(\boldsymbol{\Theta}(1))| < \frac{\delta_2}{2}\ \forall k,s \mid \boldsymbol{\Theta}(1) = \hat{\boldsymbol{\theta}}(1)) \tag{58}$$

The value of T_2 can be selected the same way T_1 was selected. In particular, the event

$$A_{2,\hat{\theta}(1)} = \{\omega \in \Omega : \lim_{t\to\infty} v_{k,s}(2,t,\omega) = a_{k,s}(\hat{\boldsymbol{\theta}}(1))\ \forall k,s\}, \tag{59}$$

occurs with probability 1, given that $\boldsymbol{\Theta}(1) = \hat{\boldsymbol{\theta}}(1)$. For every $\hat{\boldsymbol{\theta}}(1) \in \Omega_\theta$ and $\omega \in A_{2,\hat{\theta}(1)}$, there exists a $T(\hat{\boldsymbol{\theta}}(1),\omega)$ such that, given that $\boldsymbol{\Theta}(1) = \hat{\boldsymbol{\theta}}(1)$,

$$|v_{k,s}(2,t,\omega) - a_{k,s}(\hat{\boldsymbol{\theta}}(1))| < \frac{\delta_2}{2}\ \forall t \geq T(\hat{\boldsymbol{\theta}}(1),\omega), \forall k,s \tag{60}$$

Letting $T_2 = \sup_{\hat{\theta}(1)\in\Omega_\theta} \{ \sup_{\omega\in A_{2,\hat{\theta}(1)}} \{T(\boldsymbol{\theta}(0),\omega)\}\}$ gives

$$A_{2,\hat{\theta}(1)} \subseteq \{\omega \in \Omega : |v_{k,s}(2,t,\omega) - a_{k,s}(\hat{\boldsymbol{\theta}}(1))| < \frac{\delta_2}{2}\ \forall t \geq T_2, \forall k, s\} \tag{61}$$

for any $\hat{\boldsymbol{\theta}}(1) \in \Omega_\theta$. Then for any $\hat{\boldsymbol{\theta}}(1)$

$$P(|v_{k,s}(2,T_2) - a_{k,s}(\hat{\boldsymbol{\theta}}(1))| < \frac{\delta_2}{2}\forall k, s|\boldsymbol{\Theta}(1) = \hat{\boldsymbol{\theta}}(1)) \geq P(A_{2,\hat{\theta}(1)}|\boldsymbol{\Theta}(1) = \hat{\boldsymbol{\theta}}(1)) = 1 \tag{62}$$

which, using Equation 57, leads to

$$P(|v_{k,s}(2,T_2) - a_{k,s}(\boldsymbol{\Theta}(1))| < \frac{\delta_2}{2}\ \forall k, s \mid |a_{k,s}(\boldsymbol{\Theta}(1)) - a_{k,s}(\boldsymbol{\theta}(1))| < \frac{\delta_2}{2}\ \forall k, s)$$

$$= \int_{\Omega_\theta} f_{\boldsymbol{\Theta}(1)}(\hat{\boldsymbol{\theta}}(1)||a_{k,s}(\boldsymbol{\Theta}(1)) - a_{k,s}(\boldsymbol{\theta}(1))| < \frac{\delta_2}{2}\ \forall k, s))\, d\hat{\boldsymbol{\theta}}(1) = 1 \tag{63}$$

From Equations 55, 56, and 63, we see that

$$P(|v_{k,s}(2,T_2) - a_{k,s}(\boldsymbol{\theta}(1))| < \frac{\delta_2}{2}\ \forall k, s) = 1 \tag{64}$$

and, thus, from Equation 53

$$P(|\boldsymbol{\Theta}(2) - \boldsymbol{\theta}(2)| < \varepsilon_2) = 1 \tag{65}$$

Continuing the above procedure, it can be shown that for $1 \leq p \leq P$,

$$P(|\boldsymbol{\Theta}(p) - \boldsymbol{\theta}(p)| < \varepsilon_p) = 1 \tag{66}$$

if the $T_p, 1 \leq p \leq P$, are made large enough. Thus

$$P(|\boldsymbol{\Theta}(P) - \boldsymbol{\theta}(P)| < \frac{\varepsilon}{2}) = 1 \tag{67}$$

which means that

$$P(|\boldsymbol{\Theta}(P) - \boldsymbol{\theta}^*| < \varepsilon) \geq P(|\boldsymbol{\Theta}(P) - \boldsymbol{\theta}(P)| < \frac{\varepsilon}{2}, |\boldsymbol{\theta}(P) - \boldsymbol{\theta}^*| < \frac{\varepsilon}{2})$$

$$= P(|\boldsymbol{\Theta}(P) - \boldsymbol{\theta}(P)| < \frac{\varepsilon}{2}) = 1 \tag{68}$$

Thus, $\boldsymbol{\Theta}(P)$ can be made arbitrarily close to $\boldsymbol{\theta}^*$ with probability 1. This completes the proof.

REFERENCES

[1] H. Derin and H. Elliot, "Modeling and segmentation of noisy and textured images using Gibbs random fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, pp. 39–55, January 1987.

[2] P. A. Kelly, H. Derin, and K. D. Hart, "Adaptive segmentation of speckled images using a hierarchical random field model," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, pp. 1626–1641, October 1988.

[3] S. Lakshmanan and H. Derin, "Simultaneous parameter estimation and segmentation of Gibbs random fields using simulated annealing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 8, pp. 799–813, August 1989.

[4] J. Marroquin, S. Mitter, and T. Poggio, "Probabilistic solution of ill-posed problems in computational vision," *Journal of the American Statistical Association*, vol. 82, no. 397, pp. 76–89, March 1987.

[5] M. L. Comer and E. J. Delp, "Parameter estimation and segmentation of noisy or textured images using the EM algorithm and MPM estimation," *Proceedings of the 1994 IEEE International Conference on Image Processing*, November 1994, Austin, Texas, pp. 650–654.

[6] J. Zhang, J. W. Modestino, and D. A. Langan, "Maximum-likelihood parameter estimation for unsupervised model-based image segmentation," *IEEE Transactions on Image Processing*, vol. 3, no. 4, pp. 404–420, July 1994.

[7] T. N. Pappas, "An adaptive clustering algorithm for image segmentation," *IEEE Transactions on Signal Processing*, vol. 40, pp. 901–914, April 1992.

[8] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *Journal of the Royal Statistical Society B*, vol. 36, pp. 192–236, 1974.

[9] R. Kinderman and J. L. Snell, *Markov Random Fields and Their Applications*. Providence, Rhode Island: American Mathematical Society, 1980.

[10] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, no. 6, pp. 721–741, November 1984.

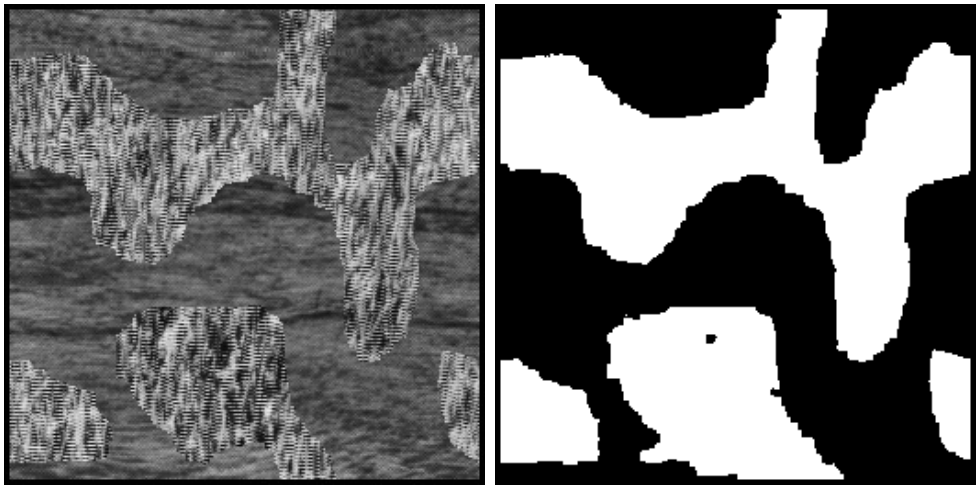
[11] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society B*, vol. 39, pp. 1–38, 1977.

[12] C. F. J. Wu, "On the convergence properties of the EM algorithm," *The Annals of Statistics*, vol. 11, no. 1, pp. 95–103, 1983.

[13] R. A. Redner and H. F. Walker, "Mixture densities, maximum likelihood and the EM algorithm," *Society for Industrial and Applied Mathematics Review*, vol. 26, no. 2, pp. 195–239, April 1984.

[14] M. Aitkin and D. B. Rubin, "Estimation and hypothesis testing in finite mixture models," *Journal of the Royal Statistical Society Series B*, vol. 47, no. 1, pp. 67–75, 1985.

[15] M. L. Comer and E. J. Delp, "Multiresolution image segmentation," *Proceedings of the 1995 IEEE International Conference on Acoustics, Speech, and Signal Processing*, May 1995, Detroit, Michigan, pp. 2415–2418.



(a) (b)

(c)

Figure 1: (a): Original image. (b): Segmented image obtained using EM/MPM. (c): Segmented image obtained using deterministic EM/MPM.

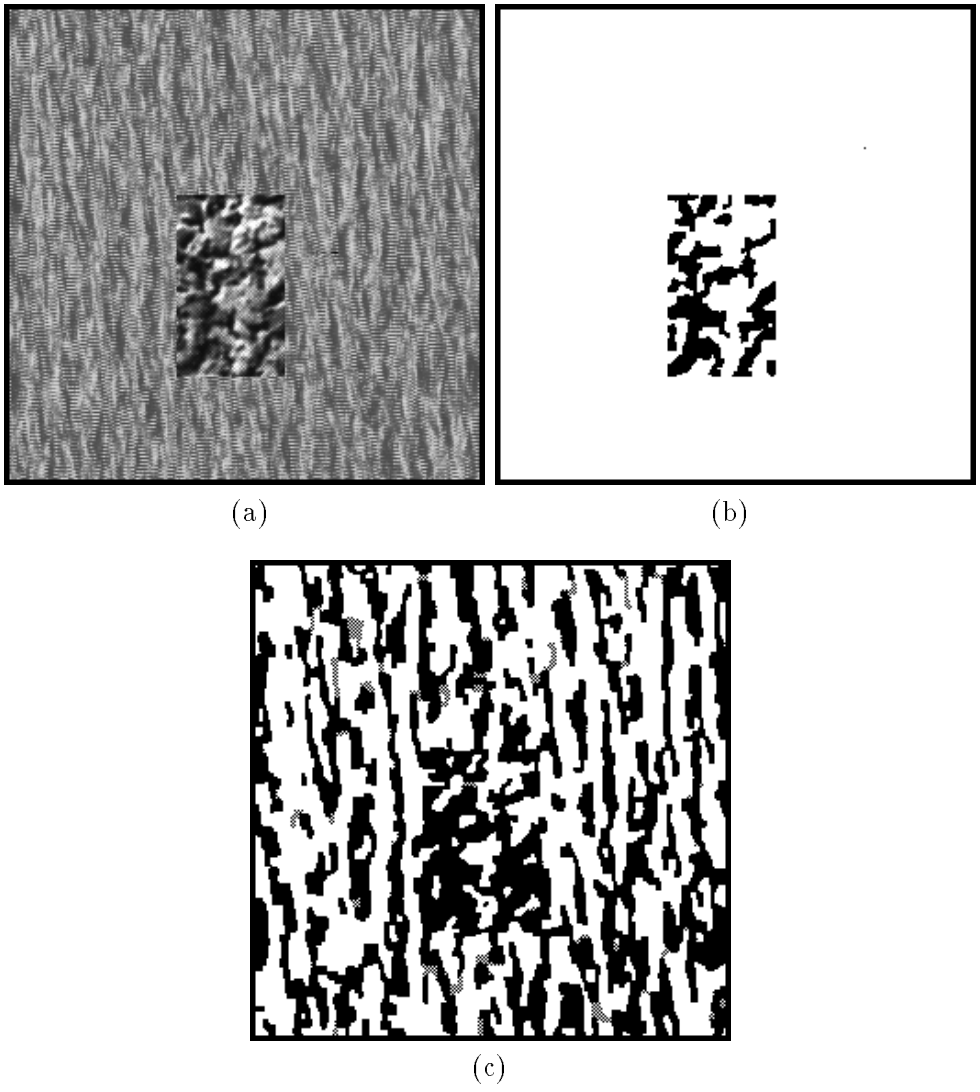


Figure 2: (a): Original image. (b): Segmented image obtained using EM/MPM. (c): Segmented image obtained using deterministic EM/MPM.

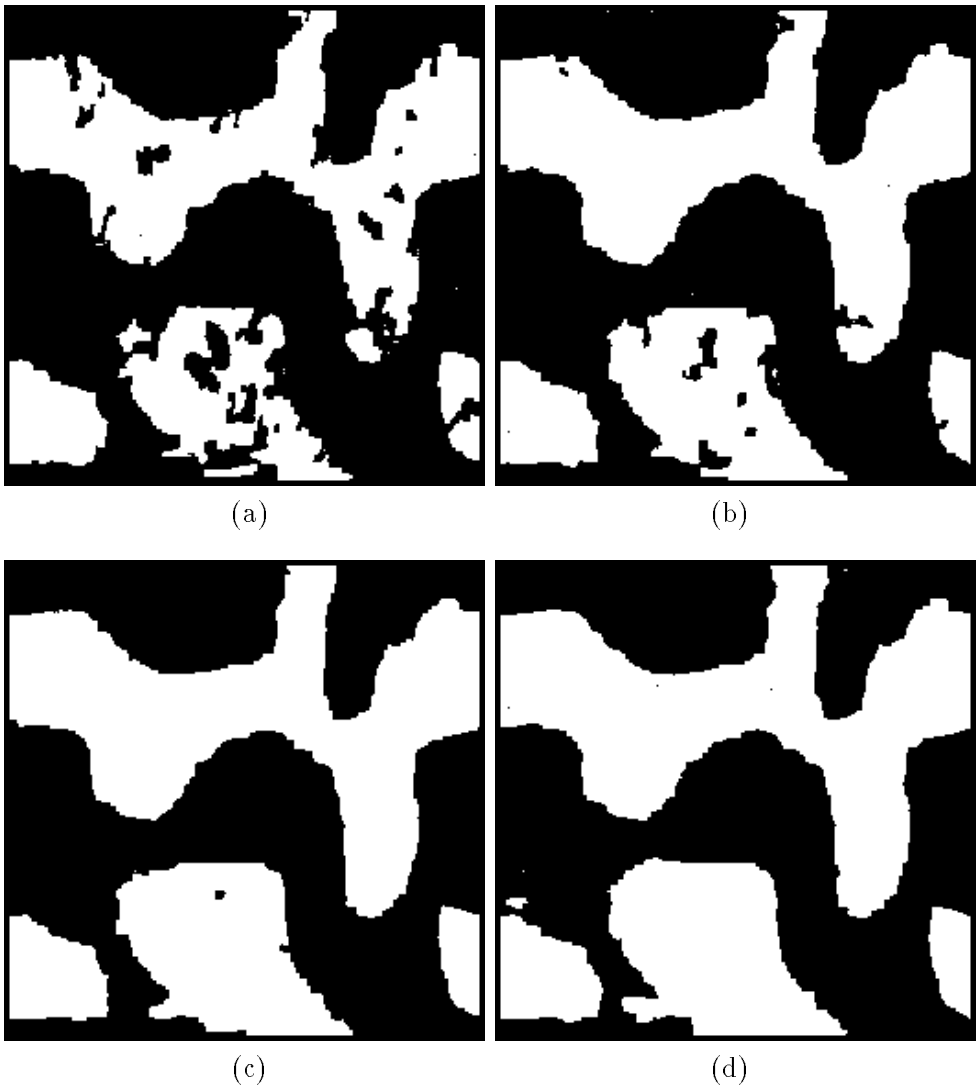
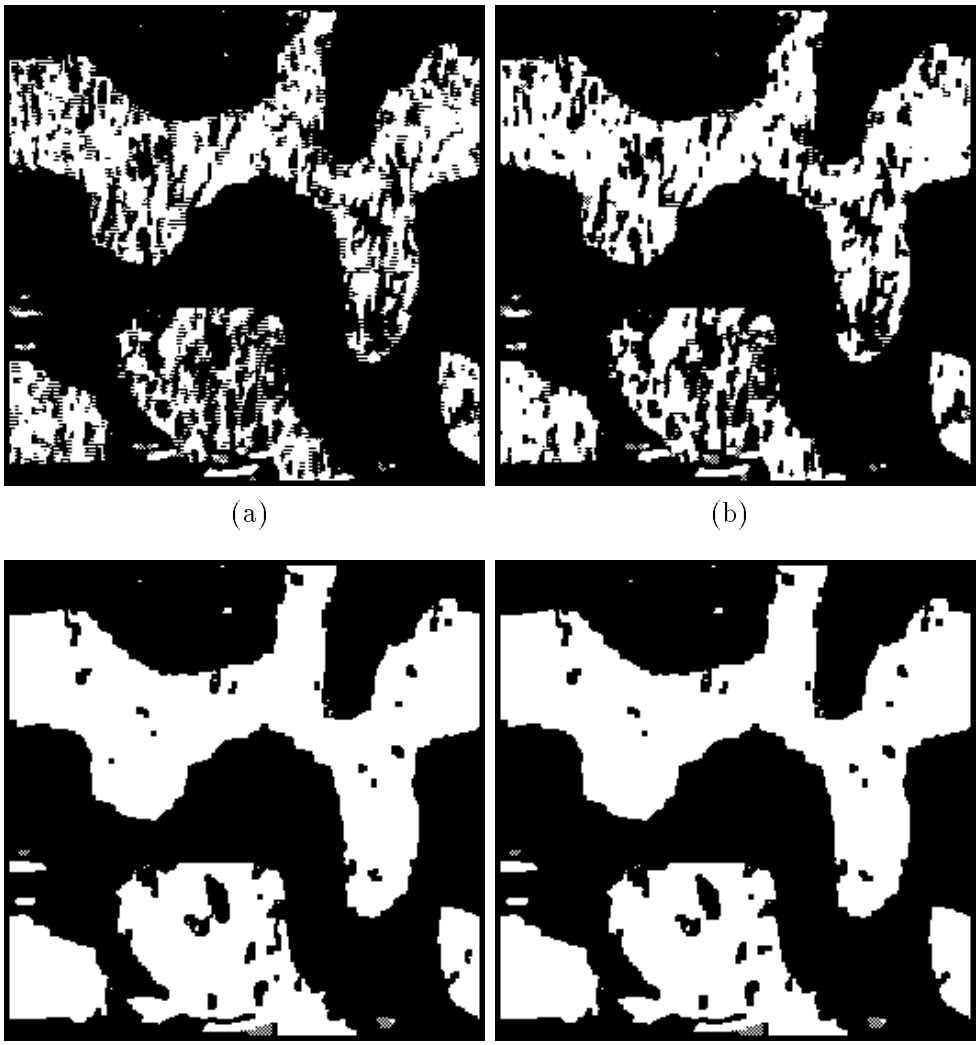


Figure 3: (a): Segmented image after 30 stages of EM/MPM. (b): Segmented image after 50 stages of EM/MPM. (c): Segmented image after 70 stages of EM/MPM. (d): Segmented image after 300 stages of EM/MPM.



(a) (b)

(c) (d)

Figure 4: (a): Segmented image after 10 iterations of deterministic EM/MPM. (b): Segmented image after 20 iterations of deterministic EM/MPM. (c): Segmented image after 500 iterations of deterministic EM/MPM. (d): Segmented image after 800 iterations of deterministic EM/MPM.

(a) (b) (c)

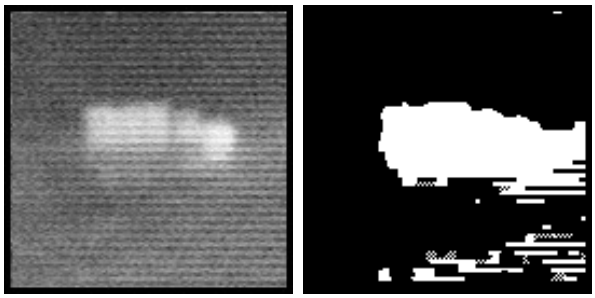
Figure 5: (a): Original image. (b): Segmented image obtained using EM/MPM with 2 classes. (c): Segmented image obtained using EM/MPM with 3 classes.

(a) (b) (c)

Figure 6: (a): Original image. (b): Segmented image obtained using EM/MPM with 2 classes. (c): Segmented image obtained using EM/MPM with 3 classes.

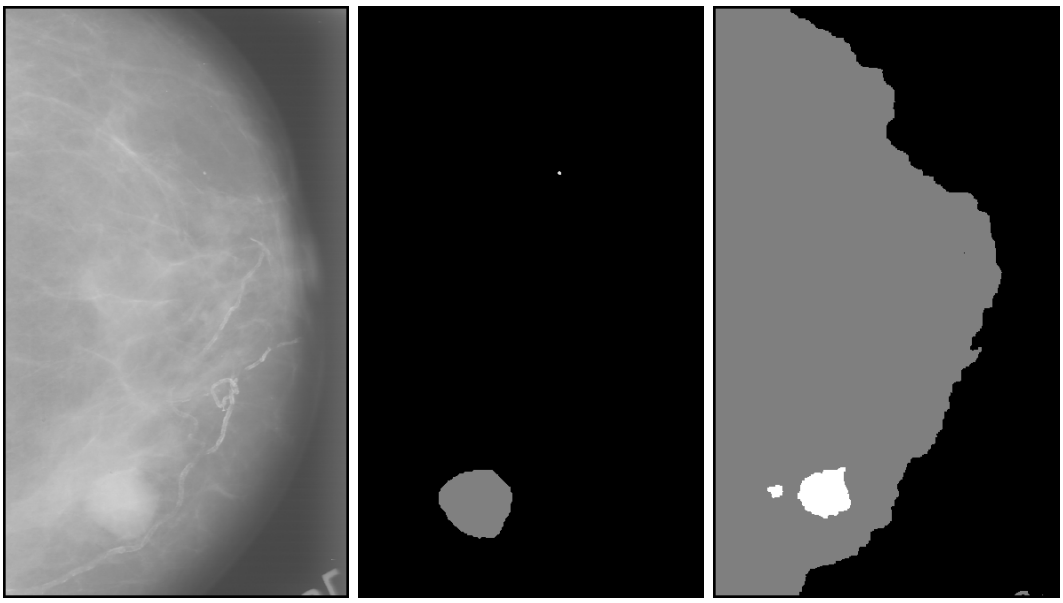
(a) (b) (c)

Figure 7: (a): Original image. (b): Segmented image obtained using EM/MPM with 2 classes. (c): Segmented image obtained using EM/MPM with 3 classes.



(a) (b)

Figure 8: (a): Original image. (b): Segmented image obtained using deterministic EM/MPM.



(a) (b) (c)

Figure 9: (a): Original image. (b): Truth image. (c): Segmented image obtained using EM/MPM.